Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 232 (2005) 179–189
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Evaluation of pathways for progression of heterogeneous breast tumors

Laura Sontag[a], David E. Axelrod[b,*]

[a]Department of Mathematics, Rutgers — The State University of New Jersey, Piscataway, NJ 08854-8019, USA
[b]Department of Genetics and the Cancer Institute of New Jersey, Rutgers — The State University of New Jersey, Piscataway, NJ 08854-8082, USA

Received 21 November 2003; received in revised form 16 July 2004; accepted 4 August 2004
Available online 29 September 2004

Abstract

To better understand the progression of heterogeneous breast cancers, four models of progession pathways have been evaluated. The models describe the progression through the grades of ductal carcinoma in situ (DCIS) 1, 2, and 3, and through the grades of invasive ductal carcinoma (IDC) 1, 2, and 3. The first three pathways, termed linear, nonlinear, and branched, describe DCIS as a progenitor of IDC, and grades of DCIS progressing into grades of IDC. The fourth pathway, termed parallel, describes DCIS and IDC as diverging from a common progenitor and progressing through grades in parallel. The best transition rates for the linear, nonlinear, and branched pathways were sought using a random search in combination with a directed search based on the Nelder–Mead simplex method. Parameter values for the parallel pathway were determined with heuristic graphs. Results of computer simulation were compared with clinically observed frequencies of grades of DCIS and grades of IDC that were reported to occur together in heterogeneous tumors. Each of the four pathways could simulate frequencies that resembled, to varying degrees, the clinical observations. The parallel pathway produced the best correspondence with clinical observations. These results quantify the traditional descriptions in which grades of DCIS are the progenitors of grades of IDC. The results also raise the alternative possibility that, in some tumors with both components, DCIS and IDC may have diverged from a common progenitor.

Keywords: Breast cancer; Tumor progression; Ductal carcinoma in situ; Invasive ductal carcinoma

1. Introduction

Diagnosis of breast cancer depends, in part, on the pathological evaluation and classification of biopsy specimens. The interpretation of the diagnostic classifications influences prognosis and therapeutic decisions. Among the classes used to describe microscopic specimens are the following: hyperplasia (increased numbers of cells), atypical hyperplasia (increased numbers of cells with abnormal morphology), ductal carcinoma in situ (increased numbers of cells with very abnormal morphology within a duct), and invasive carcinoma (abnormal cells outside of the duct). The invasive carcinomas are considered to lead to metastasis, the formation of secondary tumors, which is the most dangerous form of cancer. Ductal carcinoma in situ and invasive ductal carcinoma specimens are each further subclassified as low, intermediate, or high grade (DCIS 1, 2, or 3, or IDC 1, 2, or 3).

The proportion of patients diagnosed with DCIS, and with a mixture of DCIS and IDC, is increasing as mammography and self-examination become more common. It is important to able to predict how a tumor will progress, since this may influence treatment decisions. Several pathways describing the relationship between grades of DCIS and grades of IDC have been proposed (Buerger et al., 1999; Gupta et al., 1997; Leong et al., 2001; Mommers et al., 2001b; Roylance et al.,

*Corresponding author. Tel.: +1-732-445-2011; fax: +1-732-445-5870.
E-mail address: axelrod@biology.rutgers.edu (D.E. Axelrod).

0022-5193/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.jtbi.2004.08.002

1999, 2002). However, it is not clear which pathway best describes the biological relationship between grades of DCIS and IDC.

The purpose of this communication is to evaluate several possible pathways for breast tumor progression, with a focus on the relationship between grades of DCIS and IDC that are found to occur together in heterogeneous breast tumors. In a previous communication, linear and nonlinear pathways were investigated using genetic algorithms to search for transition rates that would match clinical observations (Subramanian and Axelrod, 2001). Such transition rates were not found, and it was concluded that the pathways were an inadequate description of the relationship between grades of DCIS and IDC. In this communication, the best transition rates for the Linear, nonlinear, and branched pathways were sought using a random search in combination with a directed search based on the Nelder–Mead simplex method. Rate constants were found for the linear and the nonlinear pathways, as well as for two additional pathways, branched and parallel. Three of these pathways describe grades of DCIS as progenitors to grades of IDC. On the other hand, the parallel pathway describes DCIS and IDC as diverging from a common progenitor and then each progressing through grades 1, 2, and 3. The parallel pathway most closely simulates the clinical observations.

2. Data and methods

2.1. Mathematical models of pathways

Four different biological pathways, termed linear, nonlinear, branched and parallel, were considered. The parallel pathway was modeled with heuristic graphs, and is described in the Results section. The linear, nonlinear, and branched pathways were interpreted mathematically as compartment models with forward transition rates between grades in an explicit series of coupled differential equations. The differential equations describing the concentration of each of the grades as a function of time (t) are given below for each pathway. The rate constants (k) for each pathway govern the rates of transition in and out of different grades of the tumor. For a set of rate constants, and given the initial condition that all the cells start with atypical hyperplasia (at time $t = 0$, [AH] = 1, and all the other concentrations are equal to 0), the differential equations were solved with MATLAB function ode45 (MathWorks, Inc., Natick, MA) to obtain the concentrations as a function of time.

2.1.1. Linear pathway

The linear pathway (Fig. 1), described previously, was unsuccessfully simulated by a genetic algorithm (Subramanian and Axelrod, 2001). The pathway starts with atypical hyperplasia (AH), continues consecutively through the three grades of DCIS, from DCIS 3 to IDC 1, and then through the three grades of IDC before reaching metastasis (M). The frequencies of each of the

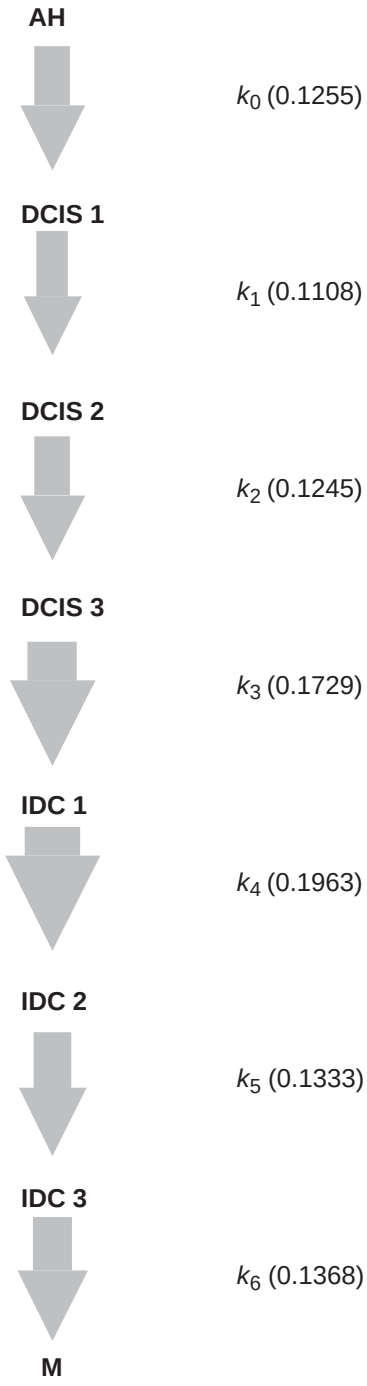

Fig. 1. Linear pathway. The rate constants shown are the average of the best fit to the Van Nuys and Holland observations, normalized to one. The thickness of each arrow is proportional to the rate constant. Atypical hyperplasia (AH), ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), and metastasis (M). Grades of DCIS and IDC are indicated by 1, 2, and 3.

grades of DCIS and IDC were observed in the specimens and reported, but the frequencies of atypical hyperplasia and of metastasis were not reported. The grades of DCIS and IDC were the subject of this study.

$$\frac{d[AH]}{dt} = -k_0[AH],$$

$$\frac{d[DCIS1]}{dt} = k_0[AH] - k_1[DCIS1],$$

$$\frac{d[DCIS2]}{dt} = k_1[DCIS1] - k_2[DCIS2],$$

$$\frac{d[DCIS3]}{dt} = k_2[DCIS2] - k_3[DCIS3],$$

$$\frac{d[IDC1]}{dt} = k_3[DCIS3] - k_4[IDC1],$$

$$\frac{d[IDC2]}{dt} = k_4[IDC1] - k_5[IDC2],$$

$$\frac{d[IDC3]}{dt} = k_5[IDC2] - k_6[IDC3].$$

2.1.2. Nonlinear pathway

The nonlinear pathway (Fig. 2) is more complete than the nonlinear pathway previously described (Subramanian and Axelrod, 2001). In the previous study, only three different values were used to describe nine possible transitions. In order to remove constraints, the present study allows different values for each of the nine transitions. The rate constant k_0 corresponds to the transition from atypical hyperplasia to the lowest grade of DCIS, k_1 and k_2 correspond to transitions between different grades of DCIS, k_3, k_4, and k_5 correspond to transitions between the grades of DCIS and IDC, k_6 and k_7 correspond to transitions between different grades of IDC, and k_8 corresponds to the transition from the highest grade of IDC to metastasis.

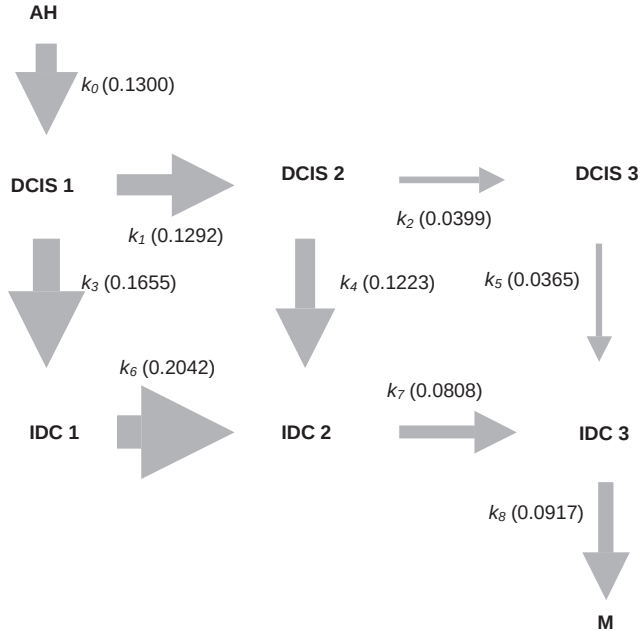

Fig. 2. Nonlinear pathway. The rate constants shown are the average of the best fit to the Van Nuys and Holland observations, normalized to one. The thickness of each arrow is proportional to the rate constant. Atypical hyperplasia (AH), ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), and metastasis (M). Grades of DCIS and IDC are indicated by 1, 2, and 3.

$$\frac{d[AH]}{dt} = -k_0[AH],$$

$$\frac{d[DCIS1]}{dt} = k_0[AH] - (k_1 + k_3)[DCIS1],$$

$$\frac{d[DCIS2]}{dt} = k_1[DCIS1] - (k_2 + k_4)[DCIS2],$$

$$\frac{d[DCIS3]}{dt} = k_2[DCIS2] - k_5[DCIS3],$$

$$\frac{d[IDC1]}{dt} = k_3[DCIS1] - k_6[IDC1],$$

$$\frac{d[IDC2]}{dt} = k_4[DCIS2] + k_6[IDC1] - k_7[IDC2],$$

$$\frac{d[IDC3]}{dt} = k_5[DCIS3] + k_7[IDC2] - k_8[IDC3].$$

2.1.3. Branched pathway

The branched pathway (Fig. 3) is a more elaborate version of the previous pathway, describing additional transitions. There are transition rate constants k_0, k_1, and k_2 from atypical hyperplasia to the three grades of DCIS, k_3 and k_4 between the grades of DCIS, k_5, k_6, and k_7 between grades of DCIS and grades of IDC, k_8 and k_9 between the grades of IDC, and k_{10}, k_{11}, and k_{12} from the grades of IDC to metastasis.

$$\frac{d[AH]}{dt} = -(k_0 + k_1 + k_2)[AH],$$

$$\frac{d[DCIS1]}{dt} = k_0[AH] - (k_3 + k_5)[DCIS1],$$

$$\frac{d[DCIS2]}{dt} = k_1[AH] + k_3[DCIS1] - (k_4 + k_6)[DCIS2],$$

$$\frac{d[DCIS3]}{dt} = k_2[AH] + k_4[DCIS2] - k_7[DCIS3],$$

$$\frac{d[IDC1]}{dt} = k_5[DCIS1] - (k_8 + k_{10})[IDC1],$$

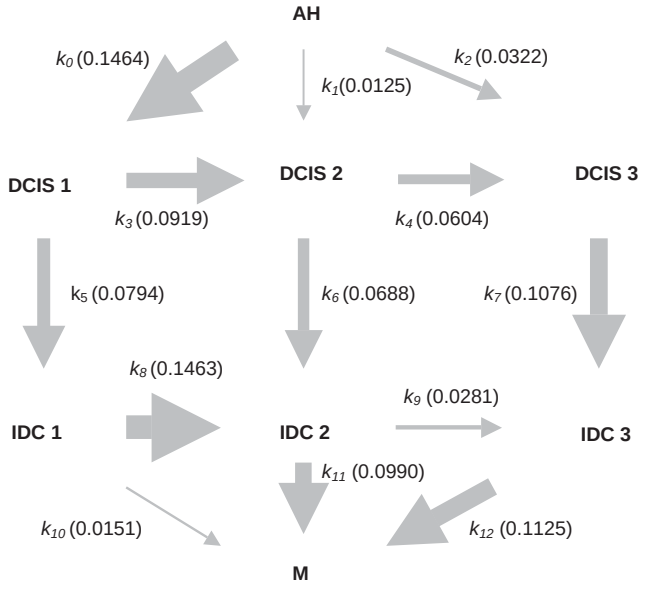

Fig. 3. Branched pathway. The rate constants shown are the average of the best fit to the Van Nuys and Holland observations, normalized to one. The thickness of each arrow is proportional to the rate constant. Atypical hyperplasia (AH), ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), and metastasis (M). Grades of DCIS and IDC are indicated by 1, 2, and 3.

$$\frac{\mathrm{d}[IDC2]}{\mathrm{d}t} = k_6[DCIS2] + k_8[IDC1] - (k_9 + k_{11})[IDC2],$$

$$\frac{\mathrm{d}[IDC3]}{\mathrm{d}t} = k_7[DCIS3] + k_9[IDC2] - k_{12}[IDC3].$$

2.2. Sources of data

Three publications were found, using the Medline literature database, which reported co-occurrence frequencies of grades of DCIS and grades of IDC in clinical specimens, and used two different systems to grade DCIS. These two systems were Van Nuys (Silverstein et al., 1995) and Holland (Holland et al., 1994). These systems each take into account nuclear and histological features, with the Van Nuys system emphasizing necrosis and the Holland system emphasizing architectural pattern. However, the two systems are common in designating three grades of DCIS. The three laboratories used the same system to grade IDC (Elston and Ellis, 1991). This system utilizes a semiquantitative evaluation of three morphological features — percent of tubule formation, degree of nuclear pleomorphism, and number of mitoses per unit area.

Data were obtained from the following publications: Gupta et al. (1997) Table 1, Holland (top), and Van Nuys (bottom); Cadman et al. (1997) Table 2, Holland, and Table 3, Van Nuys; and Leong et al. (2001) Table 2, Holland and Van Nuys. Gupta et al. reported 300 specimens by both systems; Cadman et al. reported 103 specimens by both systems; and Leong et al. reported 100 specimens by the Van Nuys system and 90 specimens by the Holland system. In order to facilitate comparisons between observations reported with the two systems, and between each system and the results simulated by the four pathways, all data reported in this communication were normalized to 498, the average of that reported with the Van Nuys (493) and with the Holland (503) systems.

The co-occurrence tables containing experimental data can be treated as contingency tables, and the

Table 1
Linear pathway. Simulated and observed co-occurrence frequencies of grades of breast ductal carcinoma in situ and grades of invasive carcinoma

IDC	DCIS		
	1	2	3
1	90.10[a], 65.66[b]	26.73, 53.54	11.88, 12.12
	94.62[c], 76.22[d]	49.80, 66.06	0, 0
2	55.45, 27.27	87.13, 117.17	55.45, 57.58
	0, 0	114.54, 127.04	74.70, 76.22
3	3.96, 4.04	25.74, 23.23	141.58, 137.38
	0, 0	0, 0	164.34, 152.45

[a]Sum of observations of Gupta, Cadman and Leong classified by the Van Nuys system, normalized to 498.
[b]Sum of observations of Gupta, Cadman and Leong classified by the Holland system, normalized to 498.
[c]Simulation by the nonlinear pathway, best fit to Van Nuys observations.
[d]Simulation by the nonlinear pathway, best fit to Holland observations.

Table 2
Nonlinear pathway. Simulated and observed co-occurrence frequencies of grades of breast ductal carcinoma in situ and grades of invasive carcinoma

IDC	DCIS		
	1	2	3
1	90.10[a], 65.66[b]	26.73, 53.54	11.88, 12.12
	60.00[c], 64.45[d]	0, 0	0, 0
2	55.45, 27.27	87.13, 117.17	55.45, 57.58
	84.00, 52.73	120.00, 140.61	78.00, 82.02
3	3.96, 4.04	25.74, 23.23	141.58, 137.38
	0, 0	0, 0	156.00, 158.19

[a]Sum of observations of Gupta, Cadman and Leong classified by the Van Nuys system, normalized to 498.
[b]Sum of observations of Gupta, Cadman and Leong classified by the Holland system, normalized to 498.
[c]Simulation by the nonlinear pathway, best fit to Van Nuys observations.
[d]Simulation by the nonlinear pathway, best fit to Holland observations.

Table 3
Branched pathway. Simulated and observed co-occurrence frequencies of grades of breast ductal carcinoma in situ and grades of invasive carcinoma

IDC	DCIS		
	1	2	3
1	90.10[a], 65.66[b]	26.73, 53.54	11.88, 12.12
	103.48[c], 90.55[d]	0, 0	0, 0
2	55.45, 27.27	87.13, 117.17	55.45, 57.58
	64.68, 51.74	103.48, 129.35	71.14, 71.14
3	3.96, 4.04	25.74, 23.23	141.58, 137.38
	0, 0	0, 0	155.22, 155.22

[a]Sum of observations of Gupta, Cadman and Leong classified by the Van Nuys system, normalized to 498.
[b]Sum of observations of Gupta, Cadman and Leong classified by the Holland system, normalized to 498.
[c]Simulation by the nonlinear pathway, best fit to Van Nuys observations.
[d]Simulation by the nonlinear pathway, best fit to Holland observations.

contingency coefficient can be calculated to test the association between the reported grades of DCIS and IDC. The contingency coefficient is a transformation of the chi-squared statistic so that the value of the contingency coefficent is in the range of 0 and 1, where high values indicate that there is dependence between variables. The contingency coefficient of the Van Nuys observations is 0.580, and of the Holland observations is 0.576.

2.3. Optimization procedure

The set of transition rate constants that produced the best correspondence of co-occurrence frequencies between each pathway and the clinically observed data was sought. A random sampling algorithm was used to obtain seed values for a directed search.

For each pathway, the following procedure was used. First, sets of rate constants were generated for the pathway using the MATLAB rand function. Rate constants were randomly chosen from a uniform distribution between zero and one. The rate constants in each set were normalized by setting their sum equal to one, to control for the speed of progression through the different grades. Each of these sets of rate constants determined the frequency of cells flowing through each of the three grades of DCIS and IDC as a function of time. The solutions to the differential equations were determined with MATLAB using the ode45 function.

The most frequently occurring grade of each DCIS and IDC was determined simultaneously at evenly spaced times within the interval in which there was more than 5% concentration of cells in DCIS and 5% in IDC. These grades are analogous to the most predominant grade of DCIS and grade of IDC in a single specimen reported by pathologists. Each pair of most frequently occurring grades was defined as a co-occurrence and was entered in the co-occurrence matrix. Sampling within this time interval is similar to the collection of multiple clinical specimens, if there is no bias in favor of collecting specimens of any particular grade.

The similarity between each generated co-occurrence matrix and the observed co-occurrence matrix was evaluated by calculating the root mean squared deviation (RMSD), as described below. For the linear and nonlinear pathways, 2000 rate constants were randomly generated and then sorted by RMSD; for the branched pathway, which has more transitions and a much larger parameter space to be explored than the other models,100,000 rate constants were randomly generated and sorted by RMSD. The set of rate constants from the random search that produced the smallest deviation, was used as the seed for the fminsearch function of the MATLAB optimization toolbox. The fminsearch function is a directed search algorithm that uses the Nelder–Mead simplex method (Lagarias et al., 1998). The co-occurrence matrices and the corresponding sets of rate constants that produced the smallest deviation are reported in the Results section.

2.4. Criteria for modeling success

The root mean squared deviation (RMSD) between the nine entries $c(i,j)$ in the clinically observed co-occurrence matrix, and the corresponding entries from the simulated co-occurrrence matrix $b(i,j)$ is:

$$\mathrm{RMSD} = \sqrt{\frac{\sum (b_{i,j} - c_{i,j})^2}{9}},$$

where, $i = 1, 2, 3$ and $j = 1, 2, 3$.

The goal was to find a minimum RMSD between the simulated results and the clinical observations, and one that is as small as the RMSD between repeated observations. The variation between repeated observations was indicated by the RMSD between observed frequencies reported by the same laboratories using two related, but different, criteria for classification, i.e. the Van Nuys system and the Holland system.

3. Results

3.1. Directed search algorithm replaces genetic algorithm

Previously, a genetic algorithm was used to search for transition rates for a linear pathway with seven rate constants, and a nonlinear pathway with three rate constants, that would reproduce the observed

co-occurrence frequencies of ductal carcinoma in situ (DCIS) grades 1, 2, and 3, and of invasive ductal carcinoma (IDC) grades 1, 2, and 3 (Subramanian and Axelrod, 2001). The genetic algorithm was trained to minimize the RMSD between the simulated results and the sum of the observations of five clinical laboratories. The best RMSDs, normalized to the sum of 1038, for the linear and the nonlinear pathways were 146 and 135, respectively. However, the best transition rates that gave the lowest RMSD produced unacceptable co-occurrence frequencies, i.e. several simulated co-occurrence frequencies were zero, including the most populated entries on the diagonal of the matrix, whereas the corresponding observed frequencies were not zero (Subramanian and Axelrod, 2001, Table 2). It was concluded that the genetic algorithm did not produce transition rates for the linear and nonlinear pathways that satisfactorily simulated the observed co-occurrence frequencies.

A revised and corrected genetic algorithm produced a different set of transition rate constants and corresponding co-occurrence matrices. The best RMSDs for the linear and nonlinear pathways were 127 and 72, respectively. Again, the transition rates that gave the best RMSDs were unacceptable in that some simulated co-occurrence frequencies were zero, but the corresponding observed frequencies were not zero. The conclusion from the results with the revised and corrected genetic algorithm is similar to that previously reported, i.e. the pathways considered and the genetic algorithm did not produce transition rates for the linear and nonlinear pathways that satisfactorily simulated the observed co-occurrence frequencies.

The inability of the genetic algorithm to find adequate transition rate constants for the linear and nonlinear pathways that could reproduce the observed co-occurrence frequencies may have been due to one or more possible circumstances. First, the genetic algorithm used may have had limitations, such as becoming trapped in a local minimum. Second, the linear and nonlinear pathways may have been inadequate because they did not fully describe the biological situation. Third, pooling five observational data sets that used several different systems of classification may have produced a combined data set that did not accurately reflect the biological situation.

Each of these possibilities has been taken into account in this communication. First, instead of using a genetic algorithm that was directed to seek transition rates that would minimize the RMSD between simulated results and observations, a sampling algorithm is used to generate a large number of random sets of transition rates. After the sets of transition rates are generated and sorted by RMSD, the one yielding the best (least) RMSD is entered as a seed into the fminsearch function of MATLAB. The fminsearch function is unlikely to become trapped in a local minimum. Second, instead of only the two pathways previously investigated, four pathways are now considered—the linear pathway, a more general nonlinear pathway with additional transition rates, a branched pathway, and a parallel pathway. Third, instead of combining data from five laboratories that used different criteria for classification, the data used comes from three clinical laboratories, each of which report classification of their same specimens by the same two systems, Van Nuys and Holland, providing a measure of reproducibility.

3.2. Linear pathway

The linear pathway allowed progression consecutively through the three grades of DCIS, from DCIS 3 to IDC 1, and then through the three grades of IDC (Fig. 1). Two thousand sets of seven transition rate constants were randomly generated by the sampling algorithm, and the resulting co-occurrence frequencies for each set were calculated. The transition rate constants were sorted based on their ability to reproduce the observed co-occurrence frequencies; that is, to produce the least RMSD between the simulated and observed co-occurrence frequencies. The set of best transition rates were used as a seed for optimization by the directed search method. The best rate constants are shown in Fig. 1. The simulated co-occurrence frequencies are shown in Table 1 together with the observed co-occurrence frequencies classified by the Van Nuys and Holland systems. Note that the linear pathway simulated four out of the nine co-occurrence frequencies with zero values, whereas none of the observed values were zero. The overall performance of the linear pathway was judged by the RMSD between the simulated co-occurrence frequencies and the observed co-occurrence frequencies (Table 5). The difference between the simulated linear pathway and the clinical observations was RMSD = 21.15. This is larger than the deviation between the observations classified by the Van Nuys system and the observations classified by the Holland system, RMSD = 18.38.

3.3. Nonlinear pathway

The nonlinear pathway previously described took into account the possibility that transitions occur between the low grade of DCIS and the low grade of IDC, between the intermediate grade of DCIS and the intermediate grade of IDC, and between the high grade of DCIS and the high grade of IDC (Subramanian and Axelrod, 2001). Previously, only three transition rate constants were allowed. Three transition rate constants may not have been adequate to describe the nine possible transitions. Therefore, a nonlinear pathway was investigated with nine different transition rate constants, one for each of the transitions. The best rate constants are shown in Fig. 2, and the simulated

co-occurrence frequencies, together with the observed co-occurrence frequencies, are shown in Table 2. The nonlinear pathway is even worse than the linear pathway in simulating the observed co-occurrence frequencies, as indicated by the RMSD = 24.54 (Table 5). In addition, the nonlinear pathway is unsatisfactory because it simulated four out of nine co-occurrence frequencies with zero values, whereas none of the observed values were zero.

3.4. Branched pathway

The branched pathway differs from the nonlinear pathway, described above, in that it includes additional transitions between atypical hyperplasia (AH) and each grade of DCIS, and between each grade of IDC and metastasis (M) (Fig. 3). This figure includes the best set of transition rate constants, i.e. those that produced the least RMSD between the simulated and observed co-occurrence frequencies. The simulated co-occurrence frequencies for the branched pathway are shown in Table 3, together with observed co-occurrence frequencies classified by the Van Nuys and Holland systems. Note that the best co-occurrence frequencies simulated by the branched pathway produced zero values for four of the nine co-occurrence frequencies, but that none of the observed values were zero. The overall performance of the branched pathway is indicated in Table 5. The difference between the simulated branched pathway and the clinical observations is RMSD = 20.62. The branched pathway produces a smaller deviation than the linear and the nonlinear pathways, but a greater deviation than that between the clinical observations, RMSD = 18.38.

3.5. Parallel pathway

Each of the pathways considered above included several implicit assumptions. It was assumed that DCIS was a progenitor of IDC and that there were one or more transitions between grades of DCIS and IDC. Also, it was assumed that all patients could be described by the same set of transition rates between grades, i.e. that there was not a group of patients that required a different description than another group of patients. The parallel pathway does not make these assumptions.

In the parallel pathway, transitions do not occur between grades of DCIS and grades of IDC. DCIS is not a progenitor of IDC, but rather, DCIS and IDC diverge from a common progenitor. There is progression from DCIS 1, to DCIS 2, to DCIS 3, and in parallel, there is progression at about the same rate from IDC 1, to IDC 2, to IDC 3 (Fig. 4). This pathway also takes into account the possibility that DCIS and IDC diverge from a common progenitor at different times in different groups of patients. Fig. 4 illustrates pairs of lineages of

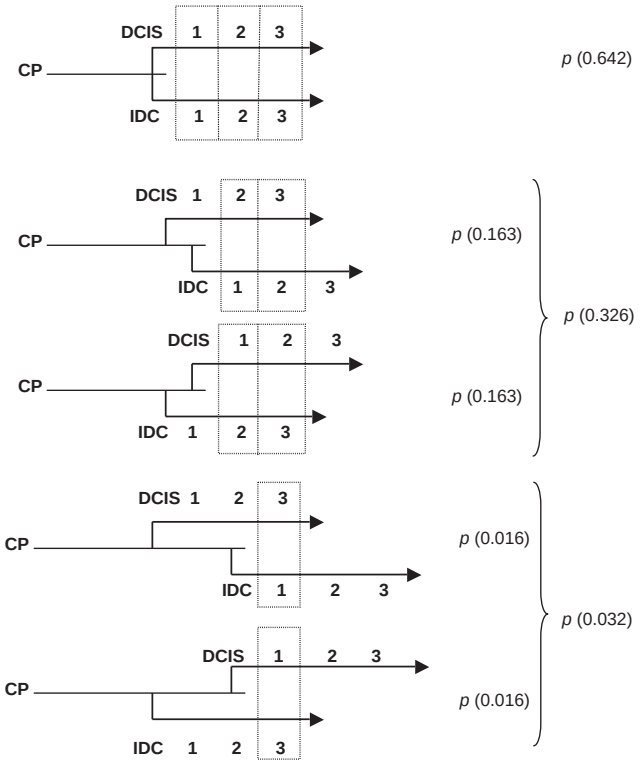

Fig. 4. Parallel pathway. DCIS and IDC diverge from a common progenitor (CP) and progress at about the same rate through grades 1, 2, and 3. Divergence may occur at the same time or at different times in different groups of patients. The proportion (p) of patients in each group is indicated. Grades that co-occur are shown above and below each other in the same box. This pathway is best at simulating the clinically observed co-occurrence frequencies.

progression, one for progression through grades of DCIS, and a separate parallel lineage for progression through grades of IDC. In some groups of patients, the lineage of DCIS starts first, and in another equal number of patients, the lineage of IDC starts first. These trajectories diverge at the same time from a common progenitor in about two-thirds of the patients (proportion, $p = 0.642$), and at different times in about one-third of the patients ($p = 0.326 + 0.032$). The proportion of patients in each of the five groups shown in Fig. 4 produce the co-occurrence frequencies in Table 4. Two-thirds of the patients have DCIS and IDC diverging at the same time, and this generates the most common co-occurrence frequencies, those on the diagonal. One-third of the patients have DCIS and IDC diverging at different times, which generates the co-occurrence frequencies in the off-diagonals. In contrast to the linear, nonlinear, and branched pathways, the parallel pathway simulates non-zero values for all nine out of nine possible co-occurrence frequencies. The overall performance of the parallel pathway, indicated by the difference between the simulated co-occurrence frequencies and the clinical observations, is RMSD = 18.82.

Table 4
Parallel pathway. Simulated and observed co-occurrence frequencies of grades of breast ductal carcinoma in situ and grades of invasive carcinoma

IDC	DCIS		
	1	2	3
1	90.10[a], 65.66[b] 106.57[c]	26.73, 53.54 40.59	11.88, 12.12 7.97
2	55.45, 27.27 40.59	87.13, 117.17 106.57	55.45, 57.58 40.59
3	3.96, 4.04 7.97	25.74, 23.23 40.59	141.58, 137.38 106.57

[a]Sum of observations of Gupta, Cadman and Leong classified by the Van Nuys system, normalized to 498.
[b]Sum of observations of Gupta, Cadman and Leong classified by the Holland system, normalized to 498.
[c]Simulated by the parallel pathway, best fit to the Van Nuys observations and to the Holland observations.

Table 5
Comparison of pathway simulations and clinical observations

Comparison		RMSD
Holland observations vs. Van Nuys observations		18.38[a]
Parallel pathway vs. Van Nuys observations Parallel pathway vs. Holland observations	17.51 20.12	18.82[b]
Branched pathway vs. Van Nuys observations Branched pathway vs. Holland observations	16.66 24.58	20.62
Linear pathway vs. Van Nuys observations Linear pathway vs. Holland observations	26.01 16.29	21.15
Nonlinear pathway vs. Van Nuys observations Nonlinear pathway vs. Holland observations	23.69 25.38	24.54

[a]Root mean squared deviation (RMSD) between clinical observations by the Van Nuys and Holland systems, normalized to 498.
[b]Average of the RMSD between pathway simulation and the two clinical observations, normalized to 498.

This is very similar to the variation between the observations by the Van Nuys system and the observations by the Holland system, RMSD = 18.38. The parallel pathway is best at simulating the clinical observations (Table 5).

4. Discussion

The breast biopsies of some patients contain two components, carcinomas in situ and invasive ductal carcinomas. Each of these components is subclassified as grade 1, 2, or 3 (well differentiated, intermediately differentiated, or poorly differentiated). In situ carcinomas are thought to be the progenitors of invasive carcinomas, but the pathway between the three grades of in situ lesions and the three grades of invasive carcinomas is not clear.

In order to determine the relationship between each of the grades of in situ and invasive carcinomas, results of simulations of four pathways were compared with clinically observed frequencies at which a grade of DCIS and a grade of IDC occur in the same biopsy. Each of the four pathways reproduced, with varying degrees of correspondence, the clinical observations. Three of these pathways describe one or more grades of carcinomas in situ progressing to one or more grades of invasive carcinomas. The pathway that best reproduces the clinically observed data is one in which grades of in situ carcinoma are not the progenitors to invasive carcinoma, but rather they have a common progenitor and there is parallel progression through the grades of in situ carcinoma and through grades of invasive carcinoma. These results supersede the previous report about two of the pathways (Subramanian and Axelrod, 2001).

The clinical observations that were used in this study reported the number of specimens that were classified as ductal carcinoma in situ (DCIS) grade 1, 2, or 3, and in addition, invasive ductal carcinoma (IDC) grade 1, 2, or 3, for a total of nine possible co-occurrence frequencies (Gupta et al., 1997; Cadman et al., 1997; Leong et al., 2001). The observed co-occurrence matrices have a pattern. The most numerous observed co-occurrence frequencies populate the diagonal, e.g. DCIS 1 and IDC 1, DCIS 2 and IDC 2, and DCIS 3 and IDC 3. This correspondence between the grades of DCIS and the grades of IDC has been noted by Gupta et al., Cadman et al., and others. Approximately two-thirds of the specimens have DCIS and IDC with corresponding grades. The remaining one-third of the specimens populate the off-diagonal positions, e.g. DCIS 1 and IDC 2, DCIS 2 and IDC 1, DCIS 2 and IDC 3, and DCIS 3 and IDC 2. The least frequent co-occurrences are far off the diagonal, e.g. DCIS 1 and IDC 3, and DCIS 3 and IDC 1. Attempts were made to reproduce the pattern of the observed co-occurrence frequencies with simulations based on four different pathways.

Some of the models investigated were motivated by pathways that other investigators described in order to summarize their observations. A simple linear pathway in which there is progression from DCIS 1, to DCIS 2, to DCIS 3, to IDC 1, to IDC 2, to IDC 3 was described by Gupta et al. and by Leong et al. However, they considered such a linear pathway unlikely, since it would not readily explain the correlation between the appearance of corresponding grades of DCIS and IDC (Gupta et al.), and the similarity of proteins expressed in corresponding grades of DCIS and IDC (Leong et al.). Nevertheless, the possibility that a linear pathway could explain the observed co-occurrence of grades of DCIS and IDC was investigated. Rate constants were found

that could produce co-occurrence frequencies similar to those observed. This simulation result could account for the similarity of cytological observations if the rate of progression through all grades was not uniform, resulting in the accumulation of corresponding grades in approximately equal amounts. In order for the linear pathway to account for the correlation of protein expression among corresponding grades of DCIS and IDC, it would be necessary to consider the possibility that the proteins that were expressed in corresponding grades were influencing the microscopic appearance of cells of corresponding grades, or less directly, that common factors influenced the appearance of cells and the expression of proteins.

Two pathways, nonlinear and branched, describe progression from grades of DCIS to corresponding grades of IDC, and in addition, progression from low grades of DCIS to higher grades of DCIS, and from low grades of IDC to higher grades of IDC. These pathways were based, in part, on reports of chromosome or protein changes in various grades of DCIS and IDC. Measuring protein expression patterns, Leong et al. (2001) and Mommers et al. (2001b) proposed that atypical hyperplasia could progress to grades DCIS, and then each grade of DCIS could progress to the corresponding grade of IDC. Roylance et al. (1999) concluded that there was not progression from well-differentiated tumors to poorly differentiated tumors, based on chromosome region losses and gains determined by comparative genomic hybridization. Their conclusion was confirmed with microsatellite markers (Roylance et al., 2002). However, based on comparative genomic hybridization, Buerger et al. (1999) proposed that atypical hyperplasia could progress to all three grades of DCIS and, in addition, that there could be progression between grades of DCIS. The nonlinear and branched pathways take into account each of these possibilities. In the nonlinear pathway, atypical hyperplasia progresses only to DCIS 1, and only IDC 3 can progress to metastasis; in the branched pathway, atypical hyperplasia can progress to all three grades of DCIS, and all three grades of IDC can progress to metastasis. Rate constants were found for both the nonlinear and branched pathways that could produce co-occurrence frequencies somewhat similar to those observed.

The linear, nonlinear, and branched pathways describe DCIS as a progenitor to IDC. This conclusion is consistent with various observations, including relative risk, gene expression (by microarrays and serial analysis of gene expression), and genome alterations (by cytogenetic analysis, comparative genomic hybridization measurements of gains and losses, and loss of heterozygosity of microsatellite markers). Some of the most distinct changes in gene expression are observed between normal epithelium and DCIS, and between atypical hyperplasia and DCIS (Porter et al., 2001; Ma et al., 2003). Many of these observations have been reviewed by van Diest (1999), Jeffrey and Pollack (2003), and Reis-Filho and Lakhani (2003). Although these observations are consistent with DCIS being a progenitor of IDC, they do not exclude the possibility that DCIS and IDC may have a common progenitor and progress in separate lineages. Observations that have been interpreted specifically as consistent with independent progression of DCIS and IDC have included measurements of nuclear morphometry (Mommers et al., 2001a; Mariuzzi et al., 2002) and microsatellite markers (Fujii et al., 1996; Lichy et al., 2000).

The possibility was also considered that DCIS is not always a progenitor to IDC, but rather, that DCIS and IDC may have a common progenitor. This model was motivated, in part, by Tsao et al. (1999, 2000) who characterized microsatellites in adjacent colorectal adenoma–carcinoma pairs. They concluded that adenomas and carcinomas diverge from a common ancestor and progress as separate lineages, rather than a sequence in which the adenomas are progenitors to carcinomas. In the parallel pathway considered in this study, the grades of breast DCIS and IDC diverge from a common progenitor and progress through their lineages separately. There is no grade of DCIS that progresses to a grade of IDC. The divergence may occur at different times in different subpopulations of patients. The parallel pathway produces simulation results that are closer to the observed co-occurrence frequencies than the other three models that describe DCIS as a progenitor to IDC.

When evaluating the results of this study, several limitations should be taken into account. Some of these are common to other modeling studies. First, the four pathways considered are not a complete set of all possible models. Some of the models were chosen because they had been proposed to explain the clinical observations, but their ability to quantitatively reproduce the clinical observations had not been investigated. Other models were elaborations of the previously proposed models or were developed for this study. Nevertheless, no attempt was made to develop and investigate a complete set of all possible models. Second, not all possible optimization procedures were used to determine the best parameter values to fit the clinical observations. A genetic algorithm, random sampling algorithm, and a directed search method were used for this study. The directed search method used initial values obtained from the random search algorithm. The direct method used here has advantages over methods such as the quasi-Newtonian method because the direct method does not depend on numerical gradients. This is an important consideration because the the error functions in this study are locally constant. Other optimization procedures could have

been used. Third, a possible change in total number of cells due to birth and death has not been described in these models.

A fourth limitation is that the evaluation of simulation results depends on the reliability of the observed data to which it is compared. The grading of clinical specimens is dependent upon human judgment, and therefore has intra- and interobserver variation. In this study, the "experimental error" of the observations is estimated by taking into account the deviations between grading of the same specimens by the same investigators as reported by each of three laboratories that used two different grading schemes. Nevertheless, since distinctions between grades 1 and 2, and between 2 and 3 are subtle it might be assumed that all specimens with DCIS and IDC should have been graded as DCIS1 and IDC 1, or DCIS 2 and IDC 2, or DCIS 3 and IDC 3. This would have resulted in entries in the co-occurrence matrix only on the diagonal. The entries reported in the off diagonal and far off diagonal would then have been the result of misgrading. It would be possible to generate the entire observed co-occurrence matrices by further assuming that the misgrading happened in the proportions shown in Fig. 4. About two-thirds of the specimens would have been graded correctly as DCIS 1 and IDC 1, or DCIS 2 and IDC 2, or DCIS 3 and IDC 3. The other one-third of the specimens would have been graded incorrectly. The proportions resulting from the differences in time of divergence from a common progenitor illustrated in Fig. 4 would be re-interpreted as probabilities of misgrading. The validity of the assumption that all specimens with both DCIS and IDC have the same grade of DCIS and IDC could be tested by using objective quantitative measurements obtained by image analysis (Axelrod et al., 2003), and establishing reproducible criteria for discriminating between grades.

DCIS accounts for 12–15% of newly diagnosed breast cancer cases, about 39,000 per year in the United States (Winchester et al., 2000; Silverstein, 2000). A better understanding of the relationship between non-invasive and invasive breast lesions could yield improvements in prevention, diagnosis, and treatment (Fisher, 1996; O'Shaughnessy et al., 2002; Lippman and Hong, 2002, Burstein et al., 2004). Although the term "ductal carcinoma in situ" has been widely used, it may be misleading. It has been proposed that it would be more appropriate to refer to the non-invasive proliferative lesions of hyperplasia, atypical hyperplasia, and ductal carcinoma in situ with the term "ductal intraepithelial neoplasia" (Tavassoli, 1998). This is more than a suggestion for a change in nomenclature, it is a change of concept. It allows reference to lesions that are abnormal without using the term "carcinoma," which implies that carcinomas in situ are obligate progenitors to invasive cancer. Our simulation results are consistent with the concept that some lesions diagnosed and referred to as carcinomas in situ may not be obligate progenitors to invasive carcinomas.

In summary, four pathways were investigated that describe possible relationships between grades of breast ductal carcinoma in situ and grades of invasive carcinoma found in the same patients. For each of the four pathways, parameter values were found that simulated co-occurrence frequencies resembling the frequencies observed clinically. The pathway that simulated the best correspondence with clinical data described DCIS and IDC as diverging from a common progenitor, and progressing through higher grades in parallel lineages. These results suggest the possibility that when breast carcinoma in situ is found with invasive carcinoma, the in situ component may not have been the progenitor of the invasive component.

Acknowledgements

We thank Dr Bala Subramanian for his foundation work and for his advice on this phase of the project, and Daniel Axelrod for suggestions about the graphic design of figures. This work was supported in part by grants from the New Jersey Commission on Cancer Research and the Busch Memorial Fund to DEA.

References

Axelrod, D.E., Chapman, J.-A., Miller, N., Christens-Barry, W., Lickley, H.L., Hanna, W., 2003. Discrimination between pathological grades of breast DCIS using image analysis. Proc. Am. Assoc. Cancer Res. 44 (2), 1267.

Buerger, H., Otterbach, F., Simon, R., Poremba, C., Diallo, R., Decker, T., Riethdorf, L., Brinkschmidt, C., Dockhorn-Dworniczak, B., Boecker, W., 1999. Comparative genomic hybridization of ductal carcinoma in situ of the breast—evidence of multiple genetic pathways. J. Pathol. 187, 396–402.

Burstein, H.J., Polyak, K., Wong, J.S., Lester, S.C., Kaelin, C.M., 2004. Ductal carcinoma in situ of the breast. N. Engl. J. Med. 350, 1430–1441.

Cadman, B.A., Ostrowski, J.L., Quinn, C.M., 1997. Invasive ductal carcinoma accompanied by ductal carcinoma in situ (DCIS): comparison of DCIS grade with grade of invasive component. The Breast 6, 132–137.

Elston, C.W., Ellis, I.O., 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow up. Histopathol 19, 403–410.

Fisher, E.R., 1996. Pathological considerations relating to the treatment of intraductal carcinoma (ductal carcinoma in situ) of the breast. CA Cancer J. Clin. 47, 52–64.

Fujii, H., Marsh, C., Cairns, P., Sidransky, D., Gabrielson, E., 1996. Genetic divergence in the clonal evolution of breast cancer. Cancer Res. 56, 1493–1497.

Gupta, S.K., Douglas-Jones, A.G., Fenn, N., Morgan, J.M., Mansel, R.E., 1997. The clinical behavior of breast carcinoma is probably determined at the preinvasive stage (ductal carcinoma in situ). Cancer 80, 1740–1745.

Holland, R., Peterse, J.L., Millis, R.R., Eusebi, V., Faverly, D., van de Vijver, M.J., Zafrani, B., 1994. Ductal carcinoma in situ: a proposal for a new classification. Semin. Diagn. Pathol. 11, 167–180.

Jeffrey, S.S., Pollack, J.R., 2003. Promise of new technologies in understanding pre-invasive breast lesions. Breast Cancer Res. 5, 320–328.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J. Optimization 9, 112–147.

Leong, A.S.-Y., Sormunen, R.T., Vinyuvat, S., Hamdani, R.W., Suthipintawong, C., 2001. Biological markers in ductal carcinoma in situ and concurrent infiltrating carcinoma. A comparison of eight contemporary grading systems. Am. J. Clin. Pathol. 115, 709–718.

Lichy, J.H., Dalbègue, F., Zavar, M., Washington, C., Tsai, M.M., Sheng, Z.-M., Taubenberger, J.K., 2000. Genetic heterogeneity in ductal carcinoma of the breast. Lab. Invest. 80, 291–301.

Lippman, S.M., Hong, W.K., 2002. Cancer prevention by delay. Clin. Cancer Res. 8, 305–313.

Ma, X.-J., Salunga, R., Tuggle, J.T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B.M., Zhou, Y.-X., Varnolt, H., Smith, B., Gadd, M., Chatfield, E., Kessler, J., Baer, T.M., Erlander, M.G., Sgroi, D.C., 2003. Gene expression profiles of human breast cancer progression. Proc. Natl. Acad. Sci. USA 100, 5974–5979.

Mariuzzi, L., Mombello, A., Granchelli, G., Rucco, V., Tarocco, E., Frank, D., Davis, J., Thompson, D., Bartels, H., Mariuzzi, G.M., Bartels, P.H., 2002. Quantitative study of breast cancer progression: Different pathways for various in situ cancers. Mod. Pathol. 15, 18–25.

Mommers, E.C.M., Poulin, N., Sangulin, J., Meijer, C.J.L.M., Baak, J.P.A., van Diest, P., 2001a. Nuclear cytometric changes in breast carcinogenesis. J. Pathol. 193, 33–39.

Mommers, E.C., Leonhart, A.M., Falix, F., Michalides, R., Meijer, C.J., Baak, J.P., Diest, P.J., 2001b. Similarity in expression of cell cycle proteins between in situ and invasive ductal breast lesions of same differentiation grade. J. Pathol. 194, 327–333.

O'Shaughnessy, J.A., Kelloff, G.J., Gordon, G.B., Dannenberg, A.J., Hong, W.J., Fabian, C.J., Sigman, C.C., Bertagnolli, M.M., Stratton, S.P., Lam, S., Nelson, W.G., Meyskens, F.L., Alberts, D.S., Follen, M., Rustgi, A.K., Papadimitrakopoulou, V., Scardino, P.T., Gazdar, A.F., Wattenberg, L.W., Sporn, M.B., Sakr, W.A., Lippman, S.M., Von Hoff, D.D., 2002. Treatment and prevention of intraepithelial neoplasia: An important target for accelerated new agent development. Clin. Cancer Res. 8, 314–346.

Porter, D.A., Krop, I.E., Nasser, S., Sgroi, D., Kaelin, C.M., Marks, J.R., Riggins, G., Polyak, K., 2001. A SAGE (Serial analysis of gene expression) view of breast tumor progression. Cancer Res. 61, 5697–5702.

Reis-Filho, J.S., Lakhani, S.R., 2003. Genetic alterations in pre-invasive lesions. Breast Cancer Res. 5, 313–319.

Roylance, R., Gorman, P., Harris, W., Liebmann, R., Barnes, D., Hanby, A., Sheer, D., 1999. Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. Cancer Res 59, 1433–1436.

Roylance, R., Gorman, P., Hanby, A., Tomlinson, I., 2002. Allelic imbalance analysis of chromosome 16q shows that grade 1 and grade III invasive ductal breast cancers follow different genetic pathways. J. Pathol. 196, 32–36.

Silverstein, M.J., 2000. Ductal carcinoma in situ of the breast. Annu. Rev. Med. 51, 17–32.

Silverstein, M.J., Poller, D.N., Waisman, J.R., Colburn, W.J., Barth, A., Gierson, E.D., Lewinsky, B., 1995. Prognostic classification of breast ductal carcinoma in situ. Lancet 345, 1154–1157.

Subramanian, B., Axelrod, D.E., 2001. Progression of heterogeneous breast tumors. J. Theor. Biol. 210, 107–119.

Tavassoli, F.A., 1998. Ductal carcinoma in situ: introduction of the concept of ductal intraepithelial neoplasia. Mod. Pathol. 11, 140–154.

Tsao, J.-L., Tavaré, S., Salovaara, R., Jass, J.R., Aaltonen, L.A., Shibata, D., 1999. Colorectal adenoma and cancer divergence: evidence of multilineage progression. Am. J. Pathol. 154, 1815–1824.

Tsao, J.-L., Yatabe, Y., Salovaara, R., Järvinen, H.J., Mecklin, J.-P., Aaltonen, L.A., Tavare, S., Shibata, D., 2000. Genetic reconstruction of individual colorectal tumor histories. Proc. Natl. Acad. Sci. USA 97, 1236–1241.

van Diest, P.J., 1999. Ductal carcinoma in situ in breast carcinogenesis. J. Pathol. 187, 383–384.

Winchester, D.P., Jeske, J.M., Goldschmidt, R.A., 2000. The diagnosis and management of ductal carcinoma in situ of the breast. CA Cancer J. Clin. 50, 184–200.